



A heat diffusion multilayer network approach for the identification of functional biomarkers in rumen methane emissions

Wang, M., Wang, H. . HY., Zheng, H., Dewhurst, R. J., & Roeheb, R. (2020). A heat diffusion multilayer network approach for the identification of functional biomarkers in rumen methane emissions. *Methods*.
<https://doi.org/10.1016/j.ymeth.2020.09.014>

[Link to publication record in Ulster University Research Portal](#)

Published in:
Methods

Publication Status:
Published online: 15/10/2020

DOI:
[10.1016/j.ymeth.2020.09.014](https://doi.org/10.1016/j.ymeth.2020.09.014)

Document Version
Author Accepted version

General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

A heat diffusion multilayer network approach for the identification of functional biomarkers in rumen methane emissions

Mengyuan Wang^a, Haiying Wang^a, Huiru Zheng^{a,1*}, Richard J. Dewhurst^b, Rainer Roehe^b

^a School of Computing, Ulster University, United Kingdom

^b Scotland's Rural College, Edinburgh, United Kingdom

Abstract

A better understanding of rumen microbial interactions is crucial for the study of rumen metabolism and methane emissions. Metagenomics-based methods can explore the relationship between microbial genes and metabolites to clarify the effect of microbial function on the host phenotype. This study investigated the rumen microbial mechanisms of methane metabolism in cattle by combining metagenomic data and network-based methods. Based on the relative abundance of 1461 rumen microbial genes and the main volatile fatty acids (VFAs), a multilayer heterogeneous network was constructed, and the functional modules associated with metabolite-microbial genes were obtained by heat diffusion. The PLS model by integrating data from VFAs and microbial genes explained 72.98% variation of methane emissions. Compared with single-layer networks, more previously reported biomarkers of methane prediction can be captured by the multilayer network. More biomarkers with the rank of top 20 topological centralities are from the PLS model of diffusion subset. The heat diffusion algorithm is different from the strategy used by the microbial metabolic system to understand methane phenotype. It inferred 24 novel biomarkers that were preferentially affected by changes in specific VFAs. Results showed that the heat diffusion multilayer network approach improved the understanding of the microbial patterns of VFA affecting methane emissions which represented by the functional genes.

Keywords: Rumen microbe, Metabolites, Metagenomics, Multilayer network, Co-occurrence network, Network diffusion algorithm

* Email address: h.zheng@ulster.ac.uk

1. Introduction

Ruminants depend on most of their nutrient supply by the fact that the complex polysaccharides of plant biomass can be degraded into monosaccharides and then fermented to produce volatile fatty acids (VFA) by the rumen microbes. VFAs are the primary source of energy absorbed from the digestive tract of ruminants. Microbial fermentation generates a surplus of hydrogen, which is the precursor of methane, as a by-product [1]. Ruminant methane emissions cause energy losses for animals and increase greenhouse gas in the environment [2]. It is estimated that the total feed energy loss due to methane emissions of cattle is as high as 12%. Methane also accounts for about one-third of agricultural methane emissions from livestock [3].

Methanogens are autotrophic microorganisms [4]. The metabolism of acetate and certain amino acids can promote the growth of methanogens [5]. The proportion and composition of diets will affect the content and composition of rumen VFA [6]. High-concentrate diets result in a high proportion of propionic acid, whilst high forage diets often generate a high proportion of acetate. The acetate type fermentation of diets with a large proportion of fibrous materials also leads to increased methane production compared to propionate type fermentation. Supplementing ingredients rich in unsaturated fatty acids to the diet can suppress the production of methane and can also change the ratio of VFAs in the rumen [6]. VFA molar composition was a significant contributor to explaining variation in methane yields from individual cattle [7]. The current rumen models are unable to accurately predict individual VFA relationships with changes in methane emission [8]. In the rumen environment, VFA and rumen microbes are interconnected and interact with each other, however, the mechanisms affecting methane emissions through changing microbial community composition and function have not been elucidated [9].

The network-based systems biology perspective may help to reveal the key relationship between the microbes and different environmental factors. Prior studies have used co-occurrence or co-abundance networks to identify different types of taxonomic and functional correlations in various microbial environments. Roehle et al. and Wallace et al. succeeded in using microbial communities or microbial genes to predict methane emissions by the combination of partial least square analysis and

network analysis [12,13]. Several studies have applied network-based approaches to reveal essential microbial genes associated with methane production [14,26,27]. By combining co-abundances networks of the microbial communities, Martínez-Álvaro et al. highlighted the importance of microbial interaction between different domains within functional niches and their changes explaining variations in methane emissions from bovine [30]. Ghanbari Maman et al. constructed the co-abundance networks in both the operational taxonomic unit and gene levels of the rumen metagenomics of sheep, which demonstrated the methane yield can be explained more precisely by used systemic views in different layers [10]. It is now clear that phenomena such as methane production are complex multi-factor traits. The interactions of different types of multi-omics data relationships will provide more complementary knowledge and additional information for the study of phenotype-related mechanisms. Multilayer networks have great potential to reveal this systematic rumen phenotypic feature [17].

Microbial interactions in the network are usually related to major changes in the structure of the entire community. Network cluster is a type of connectivity structure that forms with significantly higher frequency than random and plays important roles in complex biological networks [18]. The structure of the network is completely determined by the Laplace matrix corresponding to the network. The study found that the eigenvalues of the network Laplacian matrix can reflect some propagation dynamic characteristics of the network [19]. The heat diffusion network propagation algorithm is a node ranking and mining algorithm based on the network topology properties of the un-direct or direct graph. Theoretically, diffusion applies heat to each node in the set and lets the heat flow through connecting edges to adjacent nodes. It then produces a list of nodes ranked by the accumulated heat. The node with amount connections tends to have a higher ranking, while an isolated node tends to have a low ranking [20,21]. The Laplacian heat diffusion algorithm was applied to several studies [22-24] for the identification of latent phenotype-associated genes.

Finding the modules most relevant to phenotypic features in complex biological networks is the focus of current research. Using a combination of graph theory, network analysis, and statistical inference, complex microbial clusters can be screened to identify core biomarkers that respond to rumen methane emissions. This study constructed a multilayer network of the microbial gene-microbial gene, microbial gene-

metabolite and microbial gene-KEGG function modules to improve the insight of
95 fermentation-related mechanism of rumen methane emission. In comparison to
traditional systems biology tools, the proposed methodology has several significant
advantages. It is capable to correlate and integrate multiple data levels in a holistic
manner which may be useful to capture the heterogeneity of biological processes and
facilitate our understanding of the metabolic mechanism. As the main contributions of
100 this research, the heat diffusion algorithm was adopted as a basic searching approach
to a multilayer network to improve the effectiveness of phenotype-related biomarker
identification. The novel inferred 24 biomarkers are providing further insight into the
metabolic mechanism of rumen methane emission which is related to VFA.

2. Materials and methods

105 Based on the 7 metabolites from high-performance liquid chromatography
technology (HPLC) and relative abundance of 1461 microbial genes identified in a
metagenomics analysis, the network propagation-based approach has been developed
to infer the key functional cluster in the multilayer network. By incorporating with
partial least-squares analysis (PLS), the multilayer network was further investigated in
110 terms of network propagation algorithm.

2.1 Data Description

Data used in this study include methane emissions, the relative abundance of rumen
microbial genes, and the rumen metabolites. The study included samples of different
diets, feed additives, and breeds of beef cattle, and was balanced according to the two
115 groups with high and low feed conversion ratio [13]. Methane emissions were measured
individually for 48h in respiration chambers [25]. DNA was extracted from individual
rumen fluid samples and sequenced. To identify the microbial genes, the metagenomic
sequence reads were aligned to the KEGG (Kyoto Encyclopedia of Genes and
Genomes) genes database [12]. In total 4828 KEGG gene orthologues were identified
120 in rumen samples, of which 1461 genes showed a relative abundance of more than
0.001%. VFAs of the rumen samples were determined using HPLC as in detailed
described by the previous research [25]. The metabolites used in this study were:

Acetate, Propionate, Butyrate, ISO-butyrate, ISO-valerate, Valerate, Branched-chain
(concentrations mmol/Mol total VFA). After removing the samples with missing
125 metabolite data, a total of 34 rumen fluid samples were included in the study.

2.2 Methodology

As illustrated in Fig. 1, this study is based on the quantitative associations among
microbial genes, the interrelationships between microbial genes and metabolites, and
the known biological function information of microbial genes, combined with complex
130 network propagation algorithms to capture node association information based on edges
of different properties. The methodology predicted the phenotypic characteristics of
rumen methane emissions through tightly connected interrelationships of microbial
metabolites and microbial gene biomarkers.

2.2.1 Heterogeneous multilayer network construction

135 This study constructed three heterogeneous networks to represent different rumen
microbial functional characteristics. The co-occurrence network of microbial gene
abundance is used to represent the associations among microbial genes. Previous
studies [26] have shown that the association based on the five measures can form
different microbial gene function clusters and avoid false associations caused by the
140 nature of the metagenomic data. The simultaneous linear and non-linear relationships
between metabolites and microbial gene abundances can be represented by a similar
co-occurrence network [15]. The co-occurrence network was constructed using the
process in the previous report [27].

As shown in Table 1, the proposed method scored the associative strength by linear
145 correlations (Pearson correlation and Spearman correlation), similarity (Mutual
information) and dissimilarity (Kullback-Leibler dissimilarity and Jensen Shannon
dissimilarity) [28].

The biological associations between the microbial genes and the metabolic modules
are represented by the network of the microbial gene and functional metabolic
150 knowledge collected from the KEGG database.

The multilayer network is composed of collections of networks sharing the same nodes, but in which the edges belong to different categories or represent interactions with different natures that each interaction source has its meaning, relevance and bias [17]. In this study, the sub-networks were connected by the shared microbial gene nodes to form a heterogeneous multilayer network, which including microbial gene association, microbial gene and metabolite association and microbial gene metabolic module association.

Table 1
Summary of Quantitative Analysis Methods

Measurements	Range	Definition
Pearson correlations	[-1, +1]	$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y}$
Spearman correlations	[-1, +1]	$\rho_{rx,ry} = \frac{cov(r_x, r_y)}{\sigma_{rx} \sigma_{ry}}$
Mutual information	[0, INF]	$I(x; y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) \log \left(\frac{P(x, y)}{P(x)P(y)} \right)$
Kullback-Leibler dissimilarity	[0, INF]	$D_{KL}(x, y) = \sum P(x) \log \left(\frac{P(x)}{Q(x)} \right)$
Jensen Shannon dissimilarity	[0, INF]	$JSD(x y) = \frac{1}{2} D(x M) + \frac{1}{2} D(y M)$

^a $cov(x, y)$ is the covariance; σ_x is the standard deviation of X; σ_y is the standard deviation of Y. r_x and r_y are ranks of each observation. $P(x, y)$ is the joint probability mass function of X and Y; $P(x)$ and $P(y)$ are the marginal probability mass functions of X and Y respectively. x_i represents the Pearson correlation between the integral x and the integral i.

2.2.2 Laplacian heat diffusion algorithm

The heat diffusion algorithm is based on the calculation process of Daniel E. et al. [20]. By default, the set of selected nodes is the heat sources with each node having the same initial heat. The initial heat diffusion vector h is a column vector, in which the entries corresponding to seed nodes were set to 1, otherwise 0. At the end of diffusion, the top 90% of heat nodes were selected. The calculation is given by:

$$d = h * \exp(-Lt) \quad (1)$$

Where

- h is a vector representing the original query, and d is the result vector.
- L is the graph Laplacian defined by $D-A$.
- D is a diagonal matrix holding the degree of each node.
- A is the graph adjacency matrix of the input network.
- The scalar parameter t is the total time of diffusion, which used a default value of $t = 0.1$.

- The expression $\exp (*)$ is the matrix exponential.

2.3 Statistical analysis and tools

The PLS analysis was used to identify the most correlated microbial genes associated with methane emissions and VFA [12]. The model selection was based on the variable importance for the projection (VIP) criterion (Wold, 1995), whereby microbial parameters with a $VIP < 0.8$ contribute little to the prediction [13].

The Pearson correlation between microbial gene abundances and methane emission, PLS model and cross-validation were obtained using Matlab 2019 [29]. Correlation (Pearson, Spearman), similarity (mutual information), and dissimilarity (Euclidean distance, Kullback-Leibler dissimilarity and Jensen Shannon dissimilarity) calculations were carried out with the CoNet app [28], which was also used to do data renormalisation, multiple tests, p-value correction, bootstrapping, and Brown's p-value voting process. The correlation among metabolites and methane emissions were calculated in this study, based on the linear correlation method of Pearson and Spearman, merged p -value with Brown's method and corrected p -value by Bonferroni method (by Cytoscape. CoNet.).

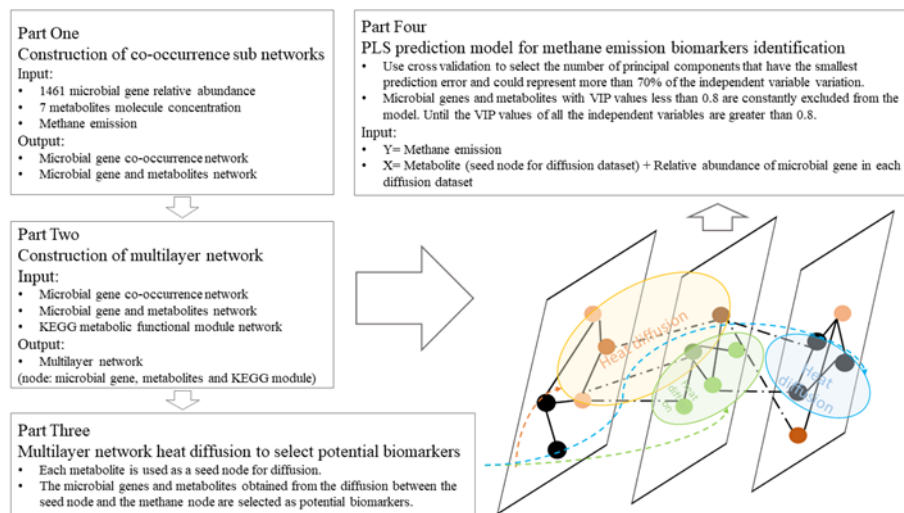


Fig. 1 Methodology pipeline

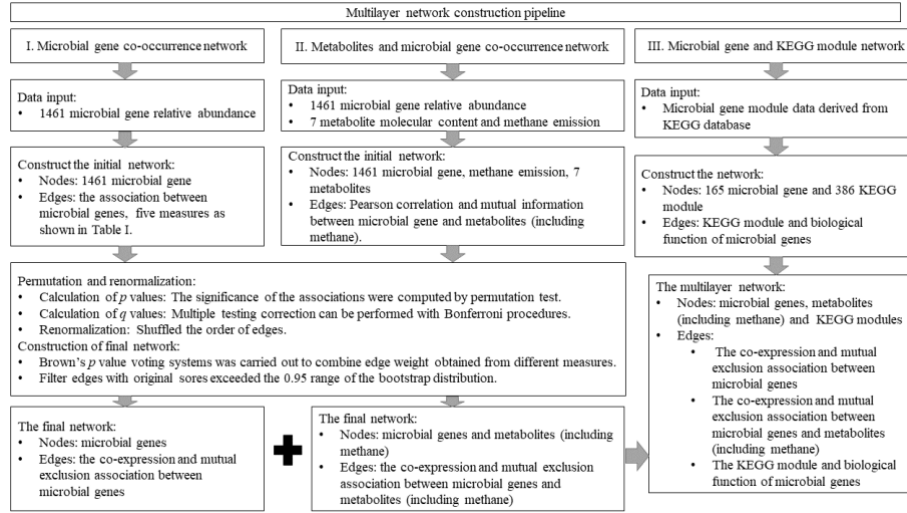


Fig. 2 Pipeline for construction of the multilayer network

3. Results

3.1 Analysis based on single-layer networks

3.1.1 Microbial Gene Co-occurrence Network

The relative abundance of 1461 microbial genes generated a microbial gene co-occurrence network with 227 nodes and 1536 edges formed 17 clusters (Fig.3). The microbial genes in 17 clusters are all co-presence correlated with each other. In the microbial gene network, 174 microbial genes have significant positive correlations with methane emissions, and only 7 microbial genes significantly negative correlated with methane emissions (Table S1). In the network, the microbial genes with the degree of the top five are K01499, K14128, K00123, K00200 and K00581. They are all involved in the methane metabolism KEGG pathway (map00680). The four microbial genes (K01499, K00123, K00200 and K00581) which encoding enzymes for methanogenesis were also reported in the similar co-abundance network research of Wang et al. [26].

3.1.2 Microbial Gene and Metabolites Co-occurrence Network

This study constructed a co-occurrence network of microbial gene and metabolite with 284 nodes and 458 edges (Fig. 4). Valerate and branched-chain fatty acid were

excluded from the network because their association with microbial genes failed the
215 threshold of multiple tests. The network showed methane emissions positive correlates
with acetate while negative correlates with propionate. The node has the highest degree
in the network is acetate, followed by propionate, which is connected to 213 and 111
microbial genes, respectively. Iso-valerate negatively correlated with K00330 and
K00340, which both encoding enzyme for NADH-quinone oxidoreductase subunit. The
220 six microbial genes (K07240, K02996, K07133, K02956, K00784 and K00788)
simultaneously associated with methane, acetate, butyrate and iso-butyrate, which
functions correspond to Ribosome, RNA transport and Thiamine metabolism.

3.1.3 Biomarkers in the single-layer network with literature evidence

Previous studies identified methane prediction biomarkers [12,13,30], in which a
225 total of 46 biomarkers were included in the network of this study. Specifically, the
microbial genes of the 2016 and 2018 studies [12,13] are mostly included in the
microbial gene co-occurrence network. The microbial genes reported in the 2020 study
are associated with acetate (Fig. 5a). There are two microbial genes (K00584 and
K14128) that are associated with acetate and iso-butyrate. Previously reported
230 biomarkers for predicting methane phenotype are connected to different metabolites in
the network. The microbial genes K00150, K01673 and K06937 are associated with
acetate and methane. Only one microbial gene (K01499) associate with acetate, iso-
butyrate and methane. There are 21 previously reported biomarker positive correlate
with acetate. The previously reported biomarker was found no association with
235 propionate or iso-valerate in this research.

There are 419 nodes and 722 edges in the microbial gene and KEGG module
network. Of the biomarkers reported in 2016 and 2018 [12,13], 24 microbial genes were
found in the KEGG module network (Fig. 5b). The microbial biomarkers reported in
2020 were not found in the KEGG module network. These methane phenotype-related
240 biomarkers come from several co-abundance networks based on microbial genes, but
in this study, no single-layer network can fully cover all biomarkers. This result
indirectly supports the necessity of an integrated heterogeneous multilayer network.

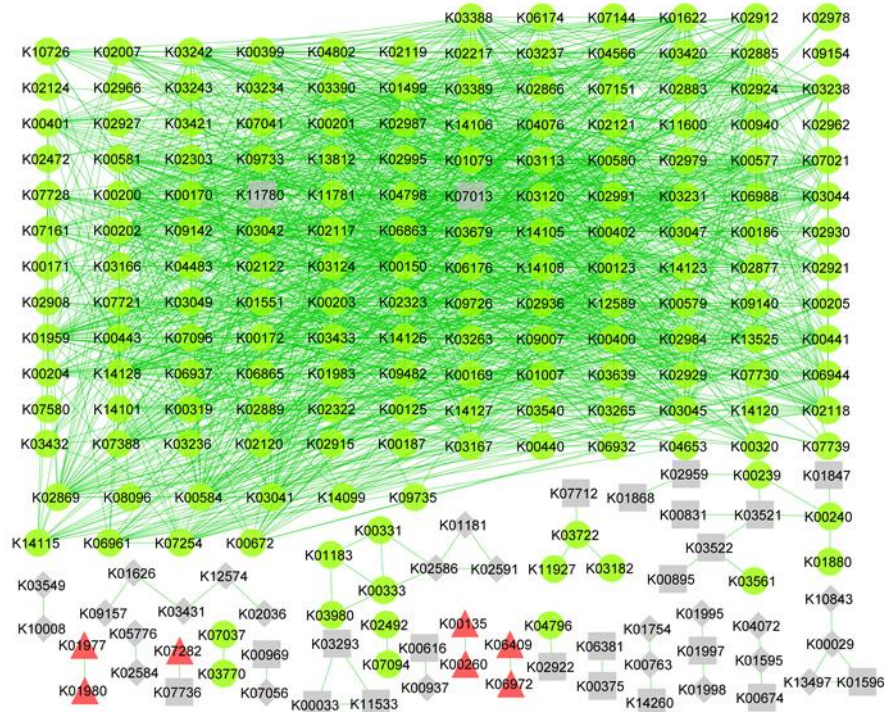
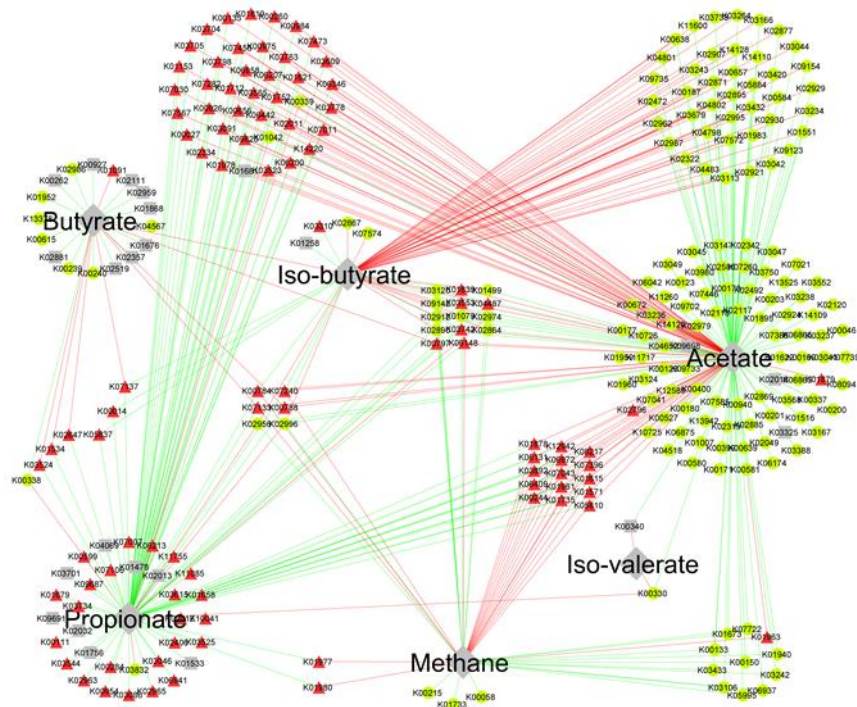


Fig. 3. Microbial gene co-occurrence network. Nodes represent microbial genes. Green circular nodes represent microbial genes significantly positive correlate with methane emissions. Red triangle nodes represent microbial genes significantly negative correlate with methane emissions. Grey nodes represent microbial genes that have no significant correlation with methane emissions. Green edges represent the co-presence relationship. Red edges represent the mutual-exclusion relationship.



250 **Fig. 4. Microbial gene and metabolite co-occurrence network.** Nodes represent microbial genes and
 255 metabolites. Green circular nodes represent microbial genes significantly positive correlate with methane
 emissions. Red triangle nodes represent microbial genes significantly negative correlate with methane
 emissions. Grey nodes represent microbial genes have no significant correlation with methane emissions.
 Green edges represent the microbial genes in the co-presence relationship. Red edges represent the microbial
 genes in the mutual-exclusion relationship.

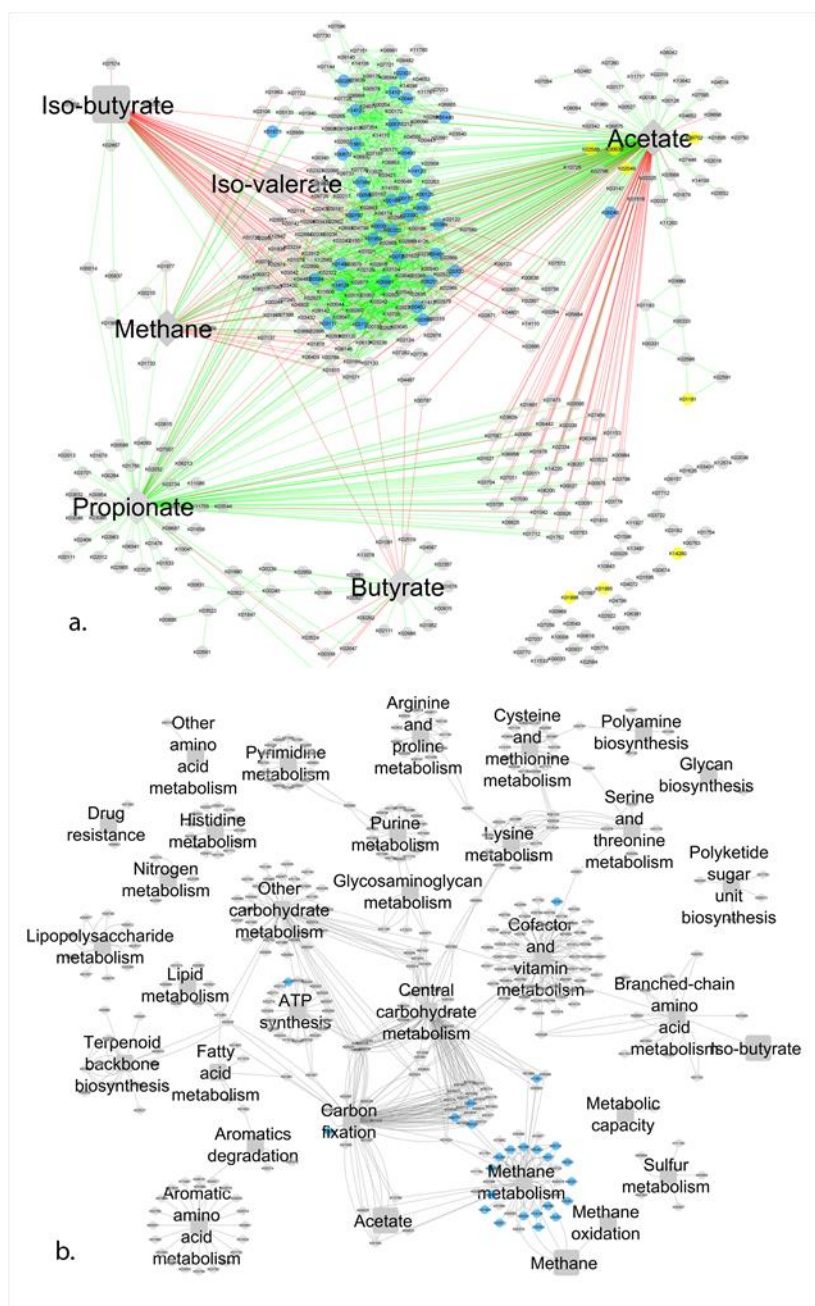


Fig. 5. Co-occurrence network and multilayer network. a. Nodes are microbial genes and metabolites. Blue nodes represent literature biomarkers of 2016 and 2018 studies. Yellow nodes represent the biomarker of the 2020 research. The green edges represent the microbial genes in the co-presence relationship, and the red edges represent the microbial genes in the mutual-exclusion relationship. b. Blue nodes represent literature biomarkers of 2016 and 2018 studies.

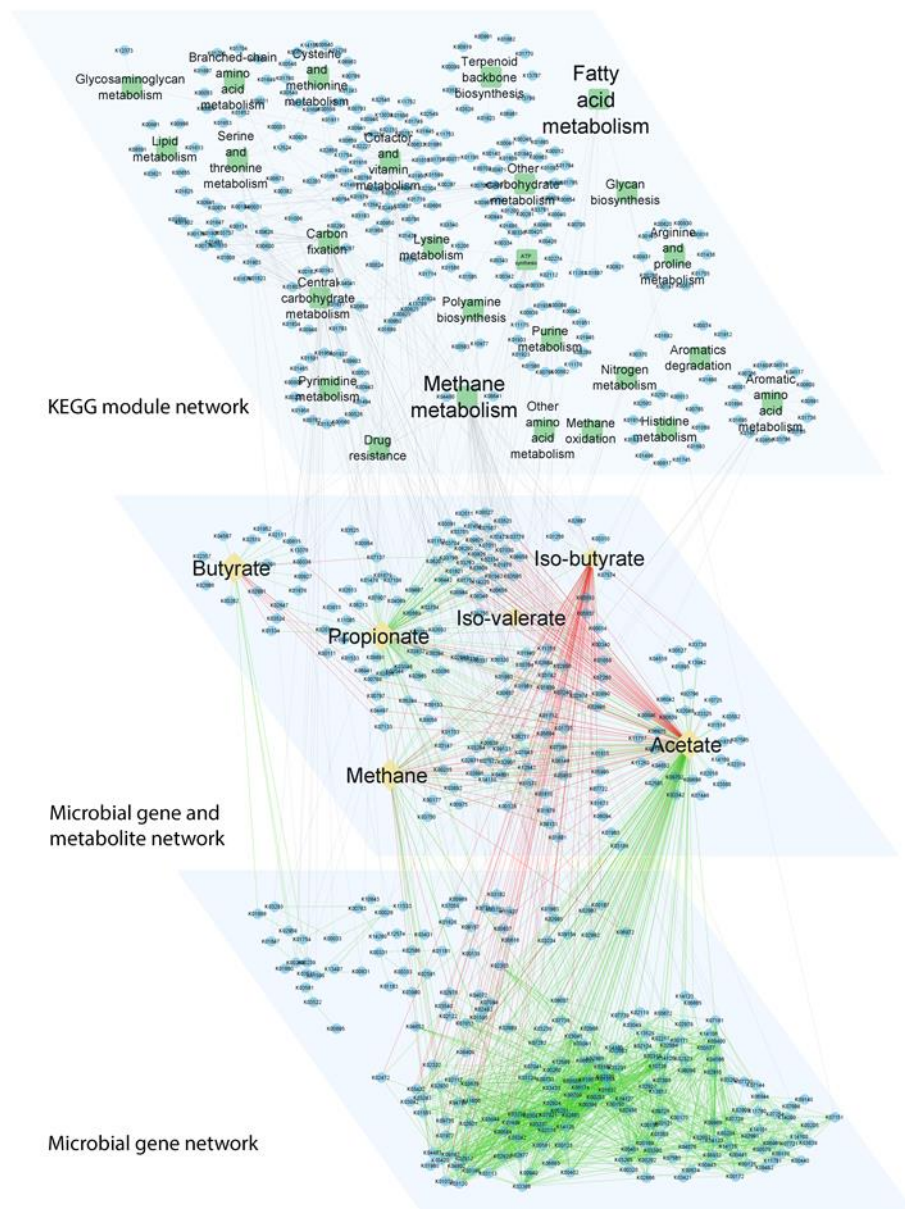
3.2 Methane emission biomarker identification based on multilayer network

3.2.1 Multilayer network

265 The microbial gene-metabolite network and microbial gene network were connected by the links between microbial genes and KEGG function modules to construct a multilayer network (Fig.7). The final network contained 663 nodes and 2418 edges. In the multilayer network, there are 99 microbial genes involved in central carbohydrate metabolism, and 97 microbial genes involved in the carbon fixation module reaction.
270 Of the 25 microbial genes shared by each layer network together, 23 involved in the methane metabolism function module. Among these three-layer networks, the microbial gene network is the densest, while the KEGG module network contained the most nodes.

3.2.2 Topological analysis of the multilayer network

275 The multilayer network was evaluated with topological parameters including clustering coefficient (0.1764), diameter (11), centralization (0.1365) and path length (4.01). The degree distribution of the multilayer network was plotted in Fig. 7 which fit the power-law distribution. Priority connectivity is the most fundamental trait for the power law of scale-free network degree distribution which means that newly joined
280 nodes have a higher probability of connecting to nodes with higher degrees. To investigate the topological relevance of each node, which may be linked to critical roles in biological mechanisms, the centrality metrics for each node including degree, betweenness, eigenvector, bridging and closeness were computed. The top 20 nodes of each centrality in the network were listed in Table 2. The central nodes in the multilayer
285 network are acetate, methane, methane metabolism module, functional genes involved in methane biosynthesis and amino acid metabolism. 16 of the 25 microbial genes connected each layer sub-network were ranked in the top 20 in topological features (Table 2). This result indicated that the multilayer network can effectively represents the microbial components of the methane phenotype related to VFA fermentation.



290

Fig. 6. Multilayer network. Nodes in the network represent microbial genes, metabolites and KEGG module respectively. Green edges represent co-presence relationships, red edges represent mutual exclusion relationships. Grey edges indicated microbial genes are involved in the KEGG metabolic module.

295

Table 2
The Top 20 Nodes Ranked by 5 Centrality Metrics

Top 20 ranked nodes in the multilayer network				
Degree	Betweenness	Eigenvector	Bridging	Closeness
Acetate	Acetate	Acetate	K03147	Acetate
Propionate	Cofactor and vitamin metabolism	K14128	K03750	Carbon fixation
Iso-butyrate	Propionate	K01499	K04518	Methane metabolism
Cofactor and vitamin metabolism	Carbon fixation	K00584	K03525	K01681
K01499	Other carbohydrate metabolism	K00581	K00954	K00788 ^c
Central carbohydrate metabolism	Central carbohydrate metabolism	K00200	K00975	K01810
K14128	Aromatic amino acid metabolism	K02885	K00527	K00169
Carbon fixation	Methane metabolism	K00201	K00600	K00244 ^c
K00200	K00626	K03242 ^a	K01712	K00150
K00584	Iso-butyrate	K03388	K01810	K00170
K03388	Pyrimidine metabolism	K00123	K01940 ^{b,c}	K01959
K00201	Cysteine and methionine metabolism	K01622	K00788 ^c	K01499
K00123	Purine metabolism	K03044	K01491	K00400
K00581	ATP synthesis	K02877	K01938	K00171
Methane	K00940	K03045	K01735 ^{b,c}	K00200
K01622	Branched-chain amino acid metabolism	K00399	K01681	K00975
Other carbohydrate metabolism	Butyrate	K02929	K01756	K01961 ^{b,c}
K02885	K00788 ^c	K02117	K00626	K02869
K03242 ^a	K00133 ^b	K02869	K01658	K00580
Methane metabolism	K00172	K03237	K00014	K00581

* a Biomarkers based on microbial gene networks. b Biomarkers based on the microbial gene-metabolite network. c Biomarkers based on diffusion subsets.

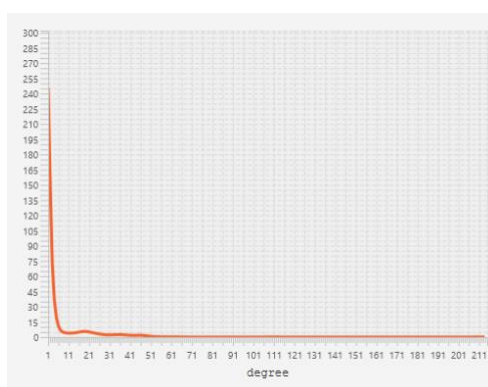


Fig. 7. Degree distribution of multilayer network. The *x*-axis represents the degree of nodes. The *y*-axis represents the number of nodes.

3.2.2 Heat diffusion on the multilayer network

According to the network propagation theory, nodes with a larger diffusion value
305 may be affected preferentially [21]. This research assumed that the network-related
nodes and corresponding interactions will change when each VFA node changing. The
heat diffusion set of each metabolite as the seed node and methane as the endpoint
represents the microbial genes or metabolites that have been preferentially affected
before the changing of each metabolite affected methane. Finally, there are 6 KEGG
310 module nodes and 280 microbial genes were captured. Only microbial gene K00788
has been included in each diffusion set.

Comparing to the impact of acetate on methane emissions, the diffusion results
indicated that acetate firstly affects the iso-butyrate. The results also showed that
changes in propionate will first affect methane among all metabolites. Before methane
315 emissions have been influenced, the changing of butyrate will first affect ATP
synthesis, carbon fixation, central carbohydrate metabolism, polyamine biosynthesis,
purine metabolism and cysteine and methionine metabolism, as well as propionate and
iso-butyrate. Iso-butyrate preferentially affects acetate before it diffuses to methane.
The propagation effect of iso-valerate on ATP synthesis, acetate and propionate are
320 much greater than that on methane emissions.

3.2.3 Biomarker identification of methane emission based on multilayer network

Fig. 8 illustrated the relationship between biomarkers of methane emission
phenotypes, diffusion sets and KEGG functional modules. Most of the biomarkers
325 reported previously have positive correlations with methane emissions, while the
biomarkers identified in this study have negative correlations with methane emissions.
All microbial genes in the microbial gene network were subjected to methane prediction
using the PLS model. The final selection of 20 microbial genes (Table S2) could explain
50.55% of methane variation. Similarly, this study used acetate, butyrate, propionate,
330 iso-butyrate and iso-valerate, as well as all microbial genes of the microbial gene and
metabolite network to predict methane emissions by the PLS model. Based on different
microbial genes and microbial gene-VFAs combinations, a PLS model with a higher

interpretation of methane prediction and fewer variables was obtained. The explained variation of methane emission by the PLS model including acetate, propionate, butyrate and 15 microbial genes can reach 72.98% (Table S2). The biomarker information is summarized in Table S3. The PLS model of propionate diffusion set can explain the methane variation up to 62.7%. The PLS model of the acetate diffusion set contains the least variables, but it can explain 60% of the methane variation, and three of the microbial genes are also included in the iso-valerate PLS model. The seven biomarkers obtained by butyrate as the seed node are distributed in 6 different subsets, and the biomarkers of the propionate heat diffusion set also involve the iso-valerate diffusion set.

Seven of all the biomarkers appeared in the top 20 rankings of topological properties (Table 2). Among them, four biomarkers from the PLS model of the microbial gene-metabolite network, and one biomarker is based on the PLS model of microbial gene network (Table 2). Among the 24 biomarkers from the PLS model of the diffusion subset, five microbial genes are in the top 20 topological centralities and three of them also in the results of the microbial metabolite-gene PLS model (Table 2). The biomarkers involved in pyruvate metabolism (K00244, K01571, K01961) were only obtained in the diffusion set using propionate as the seed node. There were two biomarkers for methane emissions involved in amino acid biosynthesis, the positively correlated one (K01940) are in the acetate and iso-valerate diffusion set, while the negatively correlated one (K01735) are included in the diffusion cluster which propionate as the seed node. Two biomarkers related to thiamine metabolism (K00788 and K04487) are in the diffusion set of butyrate. The four subunit ribosomal protein biomarkers are positively correlated with methane emissions, one (K02986) of which is in the diffusion subset of butyrate, and the other three (K02996, K02890 and K02864) are in iso-butyrate. In summary, 62 of the 88 biomarkers obtained in the literature and this study were included in the heat diffusion set of VFAs, indicating that these microbial genes were affected after the VFAs changed and eventually caused methane changes. The remaining 26 biomarkers have closely interacted with methane emissions, methane metabolism and other biomarkers.

effective measures for constructing microbial gene networks and microbial gene-metabolite network were derived based on our previous research [26,27]. The obtained multilayer network not only fully contained the methane emission biomarkers reported by other studies but also include 90% of the known methane metabolism pathway genes.

The heat diffusion algorithm is equivalent to the random walk model with restart, but it has faster calculation efficiency [16]. Some studies have shown that this large-scale network structure comprehensive algorithm can mine microbial functional associations more than the network-based node topology property analysis [19]. Besides, the network propagation algorithm based on the Laplacian matrix can obtain the ranking of nodes that have close relationships with specific nodes according to the network structure. The multilayer network of this study demonstrates the ability to identify microbial gene biomarkers for methane emissions based on abundances of the microbial genes, the concentration of the metabolites and functional KEGG pathways. The network propagation algorithm indicates the propagation path of different metabolites in the network through the network structure priority. This study used each VFA as a seed node that diffused heat to methane emissions as the end node to obtain a candidate set of potential biomarkers that represent all nodes have been affected when the seed node propagated to methane. Based on different microbial genes and microbial gene-VFAs combinations, a PLS model with a higher interpretation of methane prediction and fewer variables was obtained. In summary, 62 of the 88 biomarkers obtained in the literature and in this study were included in the heat diffusion set of VFAs, indicating that these microbial genes were affected after the VFAs were changed and eventually caused methane changes.

It has been reported that the methane phenotype and the rumen microbial community are partially heritable. Therefore, the microbial functional units clustered together within a genetic network suggests the feasibility of genetic selection for low methane emitting phenotypes. The biomarkers to predict methane productions based on rumen microbial multi-omics have complex interactions [31]. Therefore, reliable knowledge about the interactions of methane emissions and both the microbiome and metabolomics improved the identification of reliable biomarkers in this research [32,33]. Previous studies have shown VFA concentrations and ratios were significantly

correlated with methane emission, but these factors were not as robust enough for prediction. In particular, the use of microbiome information helped the development of efficient methane mitigation strategies. The improved knowledge about the rumen metabolomics is essential to explain the relevance of each domain associated with the variance of methane emissions. The systemic results of this study covered the deficiencies of previous studies by using multi-omics information in a multilayer network analysis to understand the complex interactions in biological processes and to improve the identification of reliable biomarkers associated with methane emissions.

Volatile fatty acids are the most representative rumen fermentation products, but their volatilization characteristics and wide associations make it difficult to accurately reflect the metabolic status of rumen microbes. Compared with the KEGG module information, the more detailed microbial function data (i.e. the reconstruction metabolic network) is expected to demonstrate the actual metabolic process. From the perspective of network propagation, the diffusion algorithm largely depends on the node position in the network. However, the weight and causal basis of the edge are also very valuable information and are expected to be integrated into the network in the future. By the limitation of the PLS model, when predicting the quantitative phenotype (i.e. methane emission), it should be very careful to select the variables that enter the model to avoid overfitting. In the future, we look forward to constructing a multi-level network of microbial gene abundance, microbial community information and microbial metabolomics on methane phenotypic characteristics. Another potential direction is to reveal the hidden key metabolites related to the methane phenotype through a dynamic network approach. The current study was based on the analysis of 34 rumen fluid samples with different diets and feed additive. Applying the proposed methodology to analyses a larger cohort with different diets and feed additive would be an important part of our future work.

CRedit authorship contribution statement

445 Mengyuan Wang: Conceptualization, Methodology, Implementation, Investigation,
Writing - Original Draft, Visualization.

Haiying Wang, Huiru Zheng, Richard J. Dewhurst, Rainer Roehle:
Conceptualization, Methodology, Validation, Writing - Review & Editing, Resources,
Supervision, Project administration, Funding acquisition.

450 Acknowledgement

This research is jointly supported by Ulster University and Scotland's Rural College,
U.K.

References

[1] J. L. Rombeau and S. A. Kripke, "Metabolic and Intestinal Effects of Short-Chain
455 Fatty Acids," *Journal of Parenteral and Enteral Nutrition*, vol. 14, no. 5S, pp. 181S-
185S, 1990.

[2] C. Matthews, F. Crispie, E. Lewis, M. Reid, P. W. O'Toole, and P. D. Cotter, "The
rumen microbiome: a crucial consideration when optimising milk and meat production
and nitrogen utilisation efficiency," *Gut Microbes*, vol. 10, no. 2, pp. 115–132, Mar.
460 2019.

[3] J. Wolf, G. R. Asrar, and T. O. West, "Revised methane emissions factors and
spatially distributed annual carbon fluxes for global livestock," *Carbon Balance and
Management*, vol. 12, no. 1, p. 16, Sep. 2017.

[4] J. G. Zeikus, "The biology of methanogenic bacteria.," *Bacteriol Rev*, vol. 41, no.
465 2, pp. 514–541, Jun. 1977.

[5] L. Xiao et al., "Simultaneous intensification of direct acetate cleavage and CO₂
reduction to generate methane by bioaugmentation and increased electron transfer,"
Chemical Engineering Journal, vol. 3

- 470 [6] P. Nozière, F. Glasser, and D. Sauvant, “In vivo production and molar percentages of volatile fatty acids in the rumen: a quantitative review by an empirical approach,” *animal*, vol. 5, no. 3, pp. 403–414, Mar. 2011.
- [7] J. Palarea-Albaladejo, J. A. Rooke, I. M. Nevison, and R. J. Dewhurst, “Compositional mixed modelling of methane emissions and ruminal volatile fatty acids from individual cattle and multiple experiments,” *J Anim Sci*, vol. 95, no. 6, pp. 2467–
475 2480, Jun. 2017.
- [8] H. J. van Lingen, J. G. Fadel, L. E. Moraes, A. Bannink, and J. Dijkstra, “Bayesian mechanistic modelling of thermodynamically controlled volatile fatty acid, hydrogen and methane production in the bovine rumen,” *Journal of Theoretical Biology*, vol. 480, pp. 150–165, Nov. 2019.
- 480 [9] J. Broucek, “OPTIONS TO METHANE PRODUCTION ABATEMENT IN RUMINANTS: A REVIEW,” p. 17.
- [10] L. Ghanbari Maman et al., “Co-abundance analysis reveals hidden players associated with high methane yield phenotype in sheep rumen microbiome,” *Scientific Reports*, vol. 10, no. 1, p. 4995, Mar. 2020.
- 485 [11] M. A. Jackson et al., “Detection of stable community structures within gut microbiota co-occurrence networks from different human populations,” *PeerJ*, vol. 6, p. e4303, Feb. 2018.
- [12] R. Roehe et al., “Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and
490 Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance,” *PLOS Genetics*, vol. 12, no. 2, p. e1005846, Feb. 2016.
- [13] M. D. Auffret et al., “Identification, Comparison, and Validation of Robust Rumen Microbial Biomarkers for Methane Emissions Using Diverse Bos Taurus Breeds and Basal Diets,” *Frontiers in Microbiology*, vol. 8, Jan. 2018.
- 495 [14] M. Wang, H. Wang, H. Zheng, R. J. Dewhurst, and R. Roehe, “A knowledge-driven network-based analytical framework for the identification of rumen metabolites,” *IEEE Transactions on NanoBioscience*, pp. 1–1, 2020.

- [15] R. J. Wallace et al., "The rumen microbial metagenome associated with high methane production in cattle," *BMC Genomics*, vol. 16, no. 1, Dec. 2015.
- 500 [16] Y. Yin and J. Wang, "Mechanisms of enhanced biohydrogen production from macroalgae by ferrous ion: Insights into correlations of microbes and metabolites," *Bioresource Technology*, vol. 291, p. 121808, Nov. 2019.
- [17] A. Valdeolivas et al., "Random walk with restart on multilayer and heterogeneous biological networks," *Bioinformatics*, vol. 35, no. 3, pp. 497–505, Feb. 2019.
- 505 [18] J. Fodor, M. Brand, R. J. Stones, and A. M. Buckle, "Intrinsic limitations in mainstream methods of identifying network motifs in biology," *BMC Bioinformatics*, vol. 21, no. 1, p. 165, Dec. 2020.
- [19] P. N. McGraw and M. Menzinger, "Laplacian Spectra as a Diagnostic Tool for Network Structure and Dynamics," *Phys. Rev. E*, vol. 77, no. 3, p. 031102, Mar. 2008.
- 510 [20] D. E. Carlin, B. Demchak, D. Pratt, E. Sage, and T. Ideker, "Network propagation in the Cytoscape cyberinfrastructure," *PLoS Comput Biol*, vol. 13, no. 10, Oct. 2017.
- [21] S. Lu et al., "Use of Laplacian Heat Diffusion Algorithm to Infer Novel Genes with Functions Related to Uveitis," *Front. Genet.*, vol. 9, 2018.
- [22] J. Zhang, M. Zhang, H. Zhao, and X. Xu, "Identification of proliferative diabetic retinopathy-associated genes on the protein-protein interaction network by using heat diffusion algorithm," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, p. 165794, Apr. 2020.
- 515 [23] H. Ma, H. Yang, M. R. Lyu, and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," in *Proceeding of the 17th ACM conference on Information and knowledge mining - CIKM '08*, Napa Valley, California, USA, 2008, p. 233.
- 520 [24] L. Chen, Y.-H. Zhang, Z. Zhang, T. Huang, and Y.-D. Cai, "Inferring Novel Tumor Suppressor Genes with a Protein-Protein Interaction Network and Network Diffusion Algorithms," *Molecular Therapy - Methods & Clinical Development*, vol. 10, pp. 57–67, Sep. 2018.
- 525

- [25] J. A. Rooke et al., “Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype,” *Br. J. Nutr.*, vol. 112, no. 3, pp. 398–407, Aug. 2014.
- [26] H. Wang, H. Zheng, R. J. Dewhurst, and R. Roehe, “Microbial co-presence and mutual-exclusion networks in the Bovine rumen microbiome,” in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, Nov. 2017, pp. 114–119.
- [27] H. Zheng, H. Wang, R. Dewhurst, and R. Roehe, “Improving the Inference of Co-occurrence Networks in the Bovine Rumen Microbiome,” *IEEE/ACM Trans Comput Biol Bioinform*, Nov. 2018.
- [28] K. Faust and J. Raes, “CoNet app: inference of biological association networks using Cytoscape,” *F1000Res*, vol. 5, Oct. 2016.
- [29] M. N. Triba et al., “PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters,” *Molecular BioSystems*, vol. 11, no. 1, pp. 13–19, 2015.
- [30] M. Martínez-Álvaro et al., “Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine,” *Front. Microbiol.*, vol. 11, 2020.
- [31] E. Negussie et al., “Invited review: Large-scale indirect measurements for enteric methane emissions in dairy cattle: A review of proxies and their potential for use in management and breeding decisions,” *Journal of Dairy Science*, vol. 100, no. 4, pp. 2433–2453, Apr. 2017.
- [32] C. A. McCartney, I. D. Bull, and R. J. Dewhurst, “Chemical markers for rumen methanogens and methanogenesis,” *Animal*, vol. 7, no. s2, pp. 409–417, Jun. 2013.
- [33] E. M. Ross, P. J. Moate, L. Marett, B. G. Cocks, and B. J. Hayes, “Investigating the effect of two methane-mitigating diets on the rumen microbiome using massively parallel sequencing,” *Journal of Dairy Science*, vol. 96, no. 9, pp. 6030–6046, Sep. 2013.